Some considerations concerning covariates in clinical trials

Stephen Senn

University of Glasgow | Department of Statistics

Outline

- A defence of my philosophy
  - Randomisation is valuable because it entitles you to ignore what you don't know
  - Randomisation does not entitle you to ignore what you have seen
  - This does not mean you are entitled to ignore randomisation
- The value of covariate adjustment
- Some technical issues
- Some points about effect modifiers

Three Games with Two Dice

- The object is to call the odds for getting a score of ten in rolling two dice
  - A red die and a black die
- The game is played three different ways
  - Game 1. The two die are rolled together you call the odds before they are rolled
  - Game 2. The red die is rolled you are shown the score and then call the odds before the black die is rolled
  - Game 3. You call the odds. The red die is rolled first but you are not shown it and then the black one is rolled
- How should you bet?

Table 1. Sample space for a game of chance involving two dice

	Red Die Score					
Black Die Score	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	(*10*)
5	6	7	8	9	(*10*)	11
6	7	8	9	(*10*)	11	12

Game 1: Probability = 3/36 = 1/12

Table 2. Probability of a total score of 10 given the red die score

Red Die	Probability Total = 10	Odds
1	0	0
2	0	0
3	0	0
4	1/6	1:5
5	1/6	1:5
6	1/6	1:5

Game 2: Either the probability = 0 or the probability = 1/6

Game 3: The probability = ½ x 0 + ½ x 1/6= 1/12

The Morals

- You can't treat game 2 like game 1.
  - You must condition on the information you receive in order to be wisely
  - You must use the actual data from the red die
- You can treat game 3 like game 1.
  - You can use the *distribution in probability* that the red die has
- You can't ignore an observed prognostic covariate in analysing a clinical trial just because you randomised
  - That would be to treat game 2 like game 1
- You can ignore an unobserved covariate precisely because you did randomise
  - Because you are entitled to treat game 3 like game 1

The Two Approaches

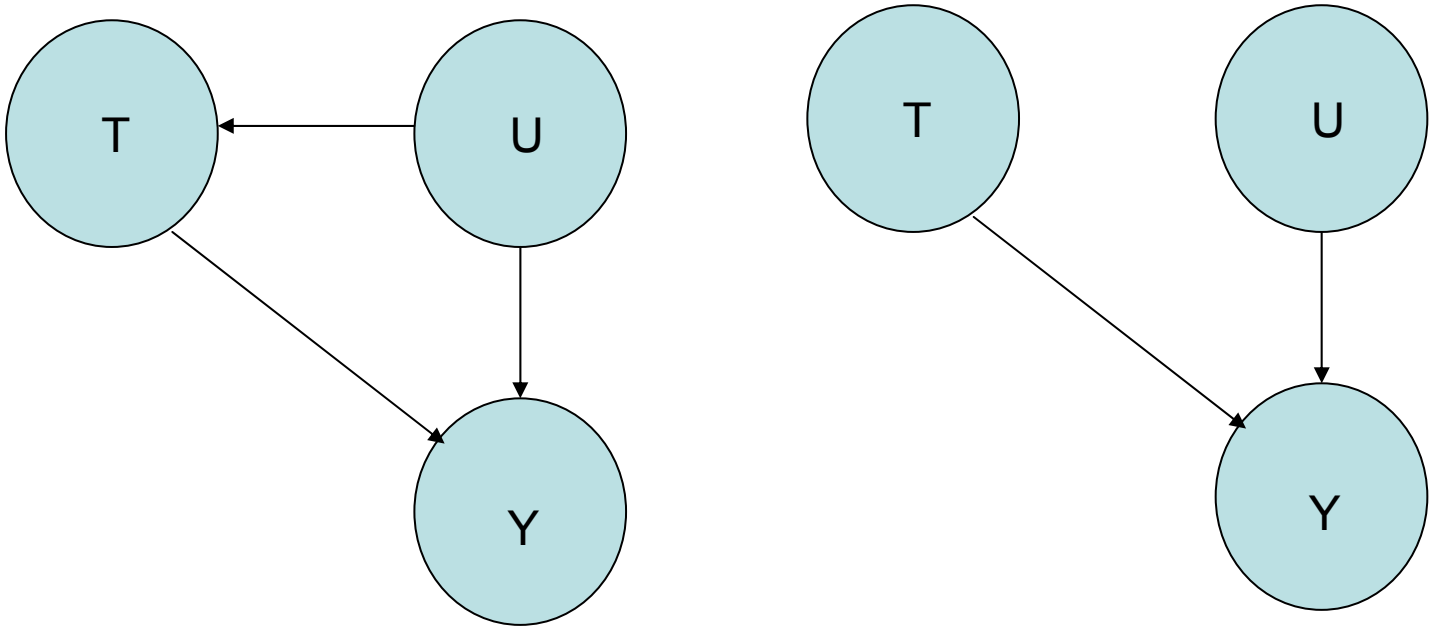
Randomisation

- Robust
- Seems to produce strong consensus for designed experiments and sampling plans
- Uses elegant symmetries to produce inference

Modelling

- Flexible
- Not limited to experiments (and random samples)
- Much more powerful in terms of scope
- Can more honestly reflect uncertainty

The Graphical Model Justification for Randomisation



See Davison, 2003, p 418

The Problem with This

	Treatment	
Grade	A	B
Moderate	44	56
Severe	56	44

- Consider three possible ways in which the observed unbalanced design on the left occurred
  - Deliberate design
  - Randomised until this pattern appeared, which was then chosen
  - Randomised and then found that by chance this pattern appeared

The Modeller's Criticism

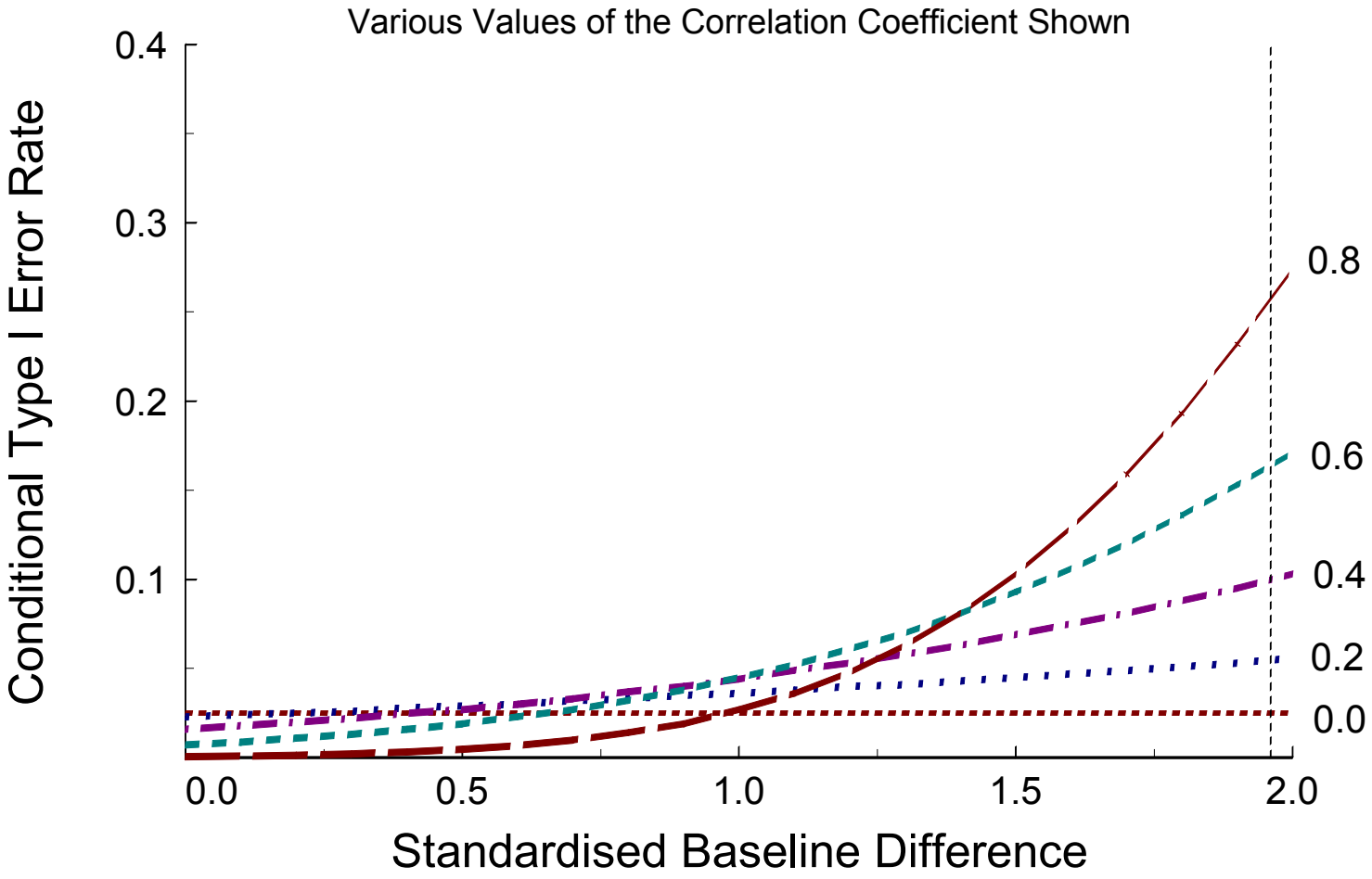
- Why should the property of the average inference over all experiments you might have run be of any relevance when making a specific inference from the experiment you did run?
- If you are at 35,000ft, four engines are on fire and the captain has had a heart-attack can you say:
  - "Why worry, on average air travel is very safe?"

A Quote from Jack Good

"The use of random sampling is a device for obtaining apparently precise objectivity but this precise objectivity is attainable, *as always*, only at the price of throwing away some information (by using a *Statistician's Stooge* who knows the random numbers but does not disclose them)…

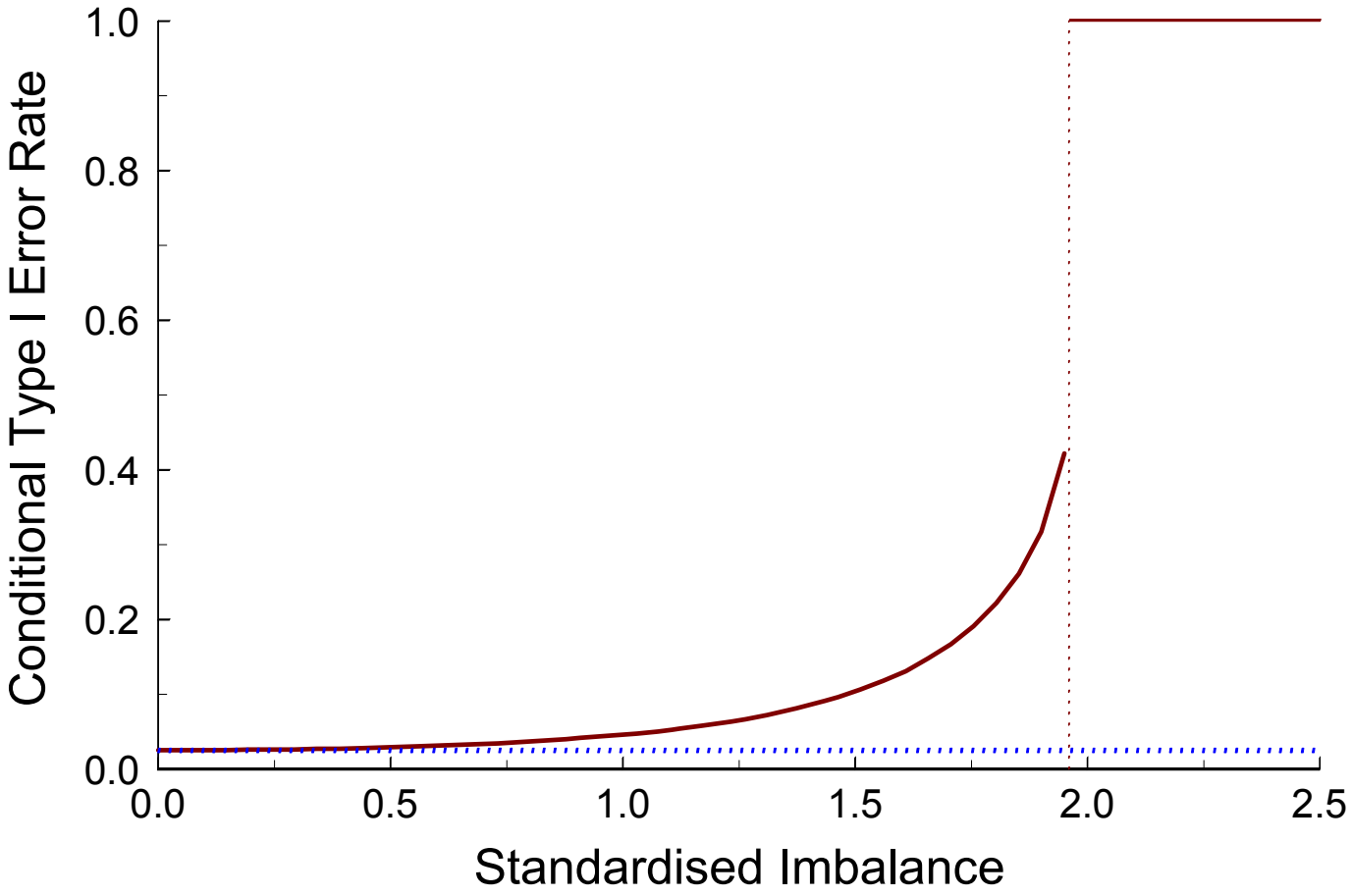
…But the use of sampling without randomization involves the pure Bayesian in such difficult judgments that, at least if he is at all Doogian, he might decide by Type II rationality, to use random sampling to save time."

Conditional Type I Error Rate for 2.5% Unconditional One Sided



Maximum Possible Conditional Type I Error Rate as a Function of Imbalance

Unconditional Type I Error Rate One-Sided is 2.5%



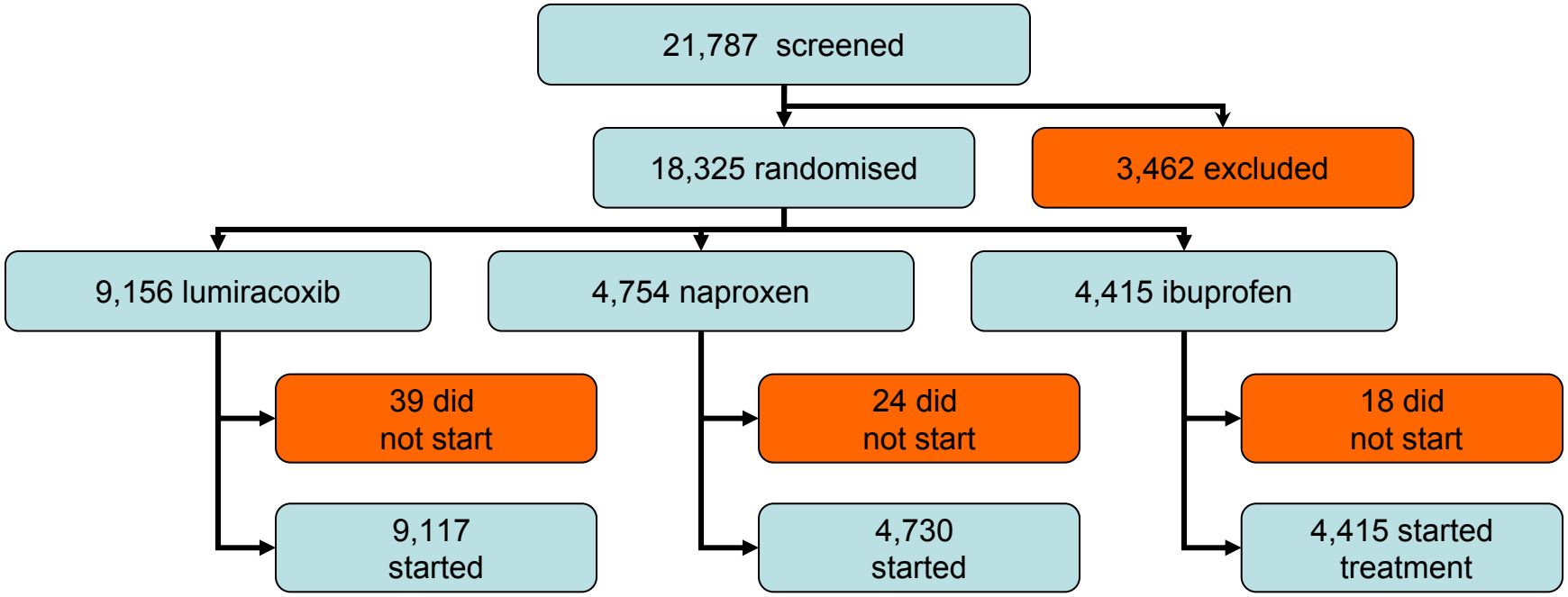
Summary of My Philosophy

- You are not entitled to ignore prognostic covariates you have observed just because you have randomised
- You are entitled to ignore covariates you have not observed because you have randomised
- For such covariates you can use their distribution in probability
- For observed covariates you must use their actual distribution
- This does *not mean* you can ignore allocation mechanisms

The TARGET study

- One of the largest studies ever run in osteoarthritis
- 18,000 patients
- Randomisation took place in two sub-studies of equal size
  - Lumiracoxib versus ibuprofen
  - Lumiracoxib versus naproxen
- Purpose to investigate CV and GI tolerability of lumiracoxib

TARGET as described in *The Lancet*



TARGET Demographics

	Sub-study I		Sub-study 2	
Demographic characteristic	**Lumiracoxib n = 4376**	**Ibuprofen n = 4397**	**Lumiracoxib n = 4741**	**Naproxen n = 4730**
Use of low-dose aspirin	975 (22%)	966 (22%)	1195 (25%)	1193 (25%)
History of vascular disease	393 (9%)	340 (8%)	588 (12%)	559(12%)
Cerebro-vascular disease	69 (2%)	65 (1%)	108 (2%)	107 (2%)
Dyslipidaemias	1030 (24%)	1025 (23%)	799 (17%)	809 (17%)

Baseline Deviances

Effect	Aspirin	Vascular History	Cerebro-vascular	Dys-lipidaemias
Substudy	23.57	70.14	13.54	117.98
Treatment-given-Substudy	0.13	5.23	0.14	0.17
Treatment	13.40	47.41	7.75	54.72

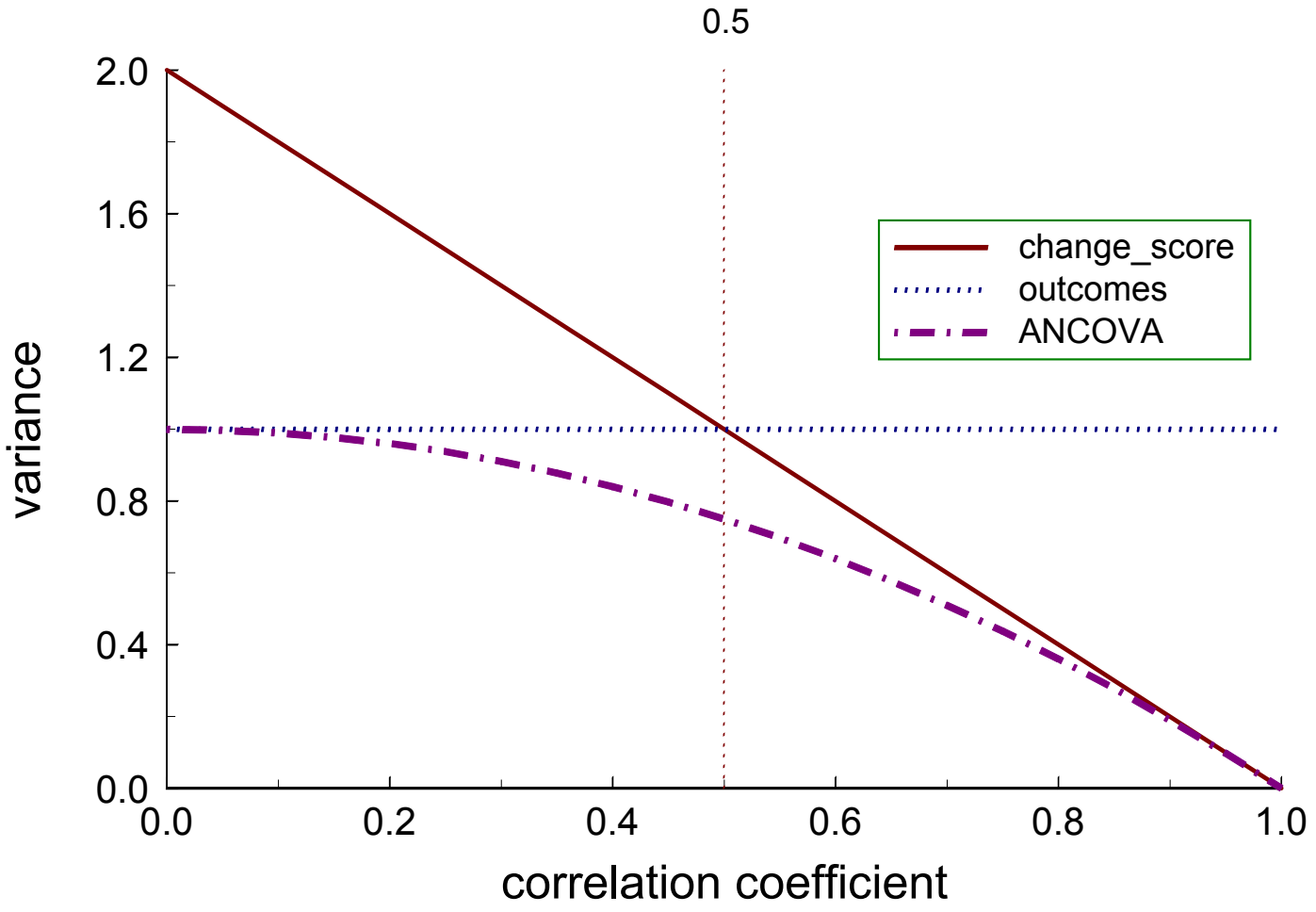
Baseline Chi-square Probabilities

Effect	Aspirin	Vascular History	Cerebro-vascular	Dys-lipidaemias
Substudy	0.0000	0.0000	0.0002	0.0000
Treatment-given-Substudy	0.9365	0.0733	0.9304	0.9194
Treatment	0.0012	0.0000	0.0208	0.0000

The Value of Covariate Adjustment

- Produces conditionally valid inferences
- May permit intelligent prediction beyond the trial
- Produces more precise inferences
- Take as an example the use of baselines….

Variances for Three Approaches to Analysis



What you learn in your first regression course

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \qquad E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad V(\hat{\boldsymbol{\beta}}) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

The Value of Balance

$$\mathrm{var}(\hat{\beta}) = (X'X)^{-1}\sigma^2$$

Variance multiplier for the treatment effect

$$= \begin{pmatrix} a_{11} & a_{12} & a_{1k} \\ a_{12} & a_{22} & \\ & & \\ a_{1k} & & a_{kk} \end{pmatrix} \sigma^2$$

The value of σ^2 depends on the model.

The value of a_{22} depends on the design and this only achieves its lower bound when covariates are balanced.

$$a_{22} \geq 2/n$$

In practice adding more covariates makes σ^2 lower but increases a_{22}.

Consequences

- In a frequentist framework there is a trade-off in fitting covariates
  - Reduction of expected residual variance
  - Increase in expected loss to non-orthogonality
- There is something not quite right about this from a philosophical point of view
- More information could be bad
  - NB Gauss-Markov theorem does not apply to stochastic regressors
- A Bayesian resolution may exist using informative priors
- There are some deep issue here

Non-linear models

- Here the situation is more complex
- Usually models over means are not the same as means over models
- My view is that you should use conditional models but you may wish to issue marginal predictions
- To do this you need a relevant distribution of covariates

In summary

- Fitting covariates is valuable
- It can provide an increase in information equivalent to having studied many more patients
- It can provide insight into sources of variation
- It can provide a basis for prediction

Some Technical Issues

- Covariate adjustment versus stratification
- Continuous predictors
- How many covariates?
- Additive models

Covariate Adjustment v Stratification

- Consider a case of a two armed trial with a binary covariate
- Stratified analysis will estimate the treatment effect within each stratum and then combine them
- Two degrees of freedom will be eliminated for each stratum
- This makes four in total
  - They correspond to intercept 1, treatment, 1, covariate, 1, interaction 1.
- ANCOVA on the other hand will eliminate 3 degrees of freedom
  - Intercept 1, treatment 1, covariate 1

Chuang-Stein and Tong

In a two group clinical trial with a binary outcome patients are also classified by a dichotomous covariate

Three statisticians analyse the same trial and produce the following estimates of

log-odds ratios (standard errors) [p-values]

depending on whether the code the covariate **0,1** or **1,0** or **-1,1**:

-1.43 (0.528) [0.007]

-0.25 (0.703) [0.727]

-0.84 (0.439)[0.057]

Chuang-Stein C, Tong D. The impact of parameterization on the interpretation of the main-effect term in the presence of an interaction. *Drug Inf J*. 1996; 30: 421-424.

-1.43 (0.528) [0.007] Stratum 1 (Coding A)

-0.25 (0.703) [0.727] Stratum 2 (Coding B)

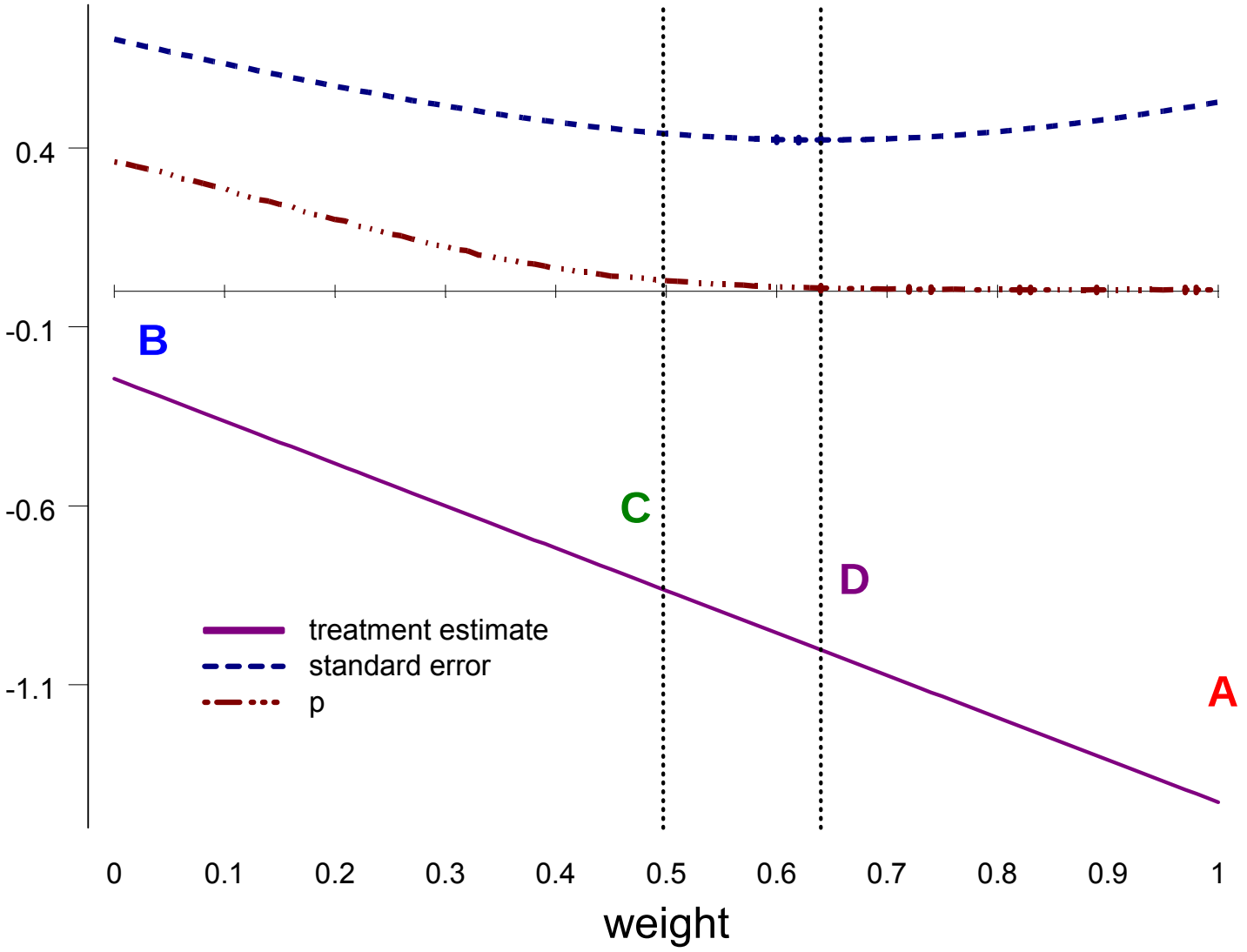
-0.84 (0.439)[0.057] Unweighted mean (Coding C)

But the meta-analyst would weight by precision. This yields:

-1.00 (0.42) [0.018]

This can be produced by logistic regression if the codes -0.36, 0.64 are used (Coding D)

Graphical Representation of Approaches to Logistic Regression Example

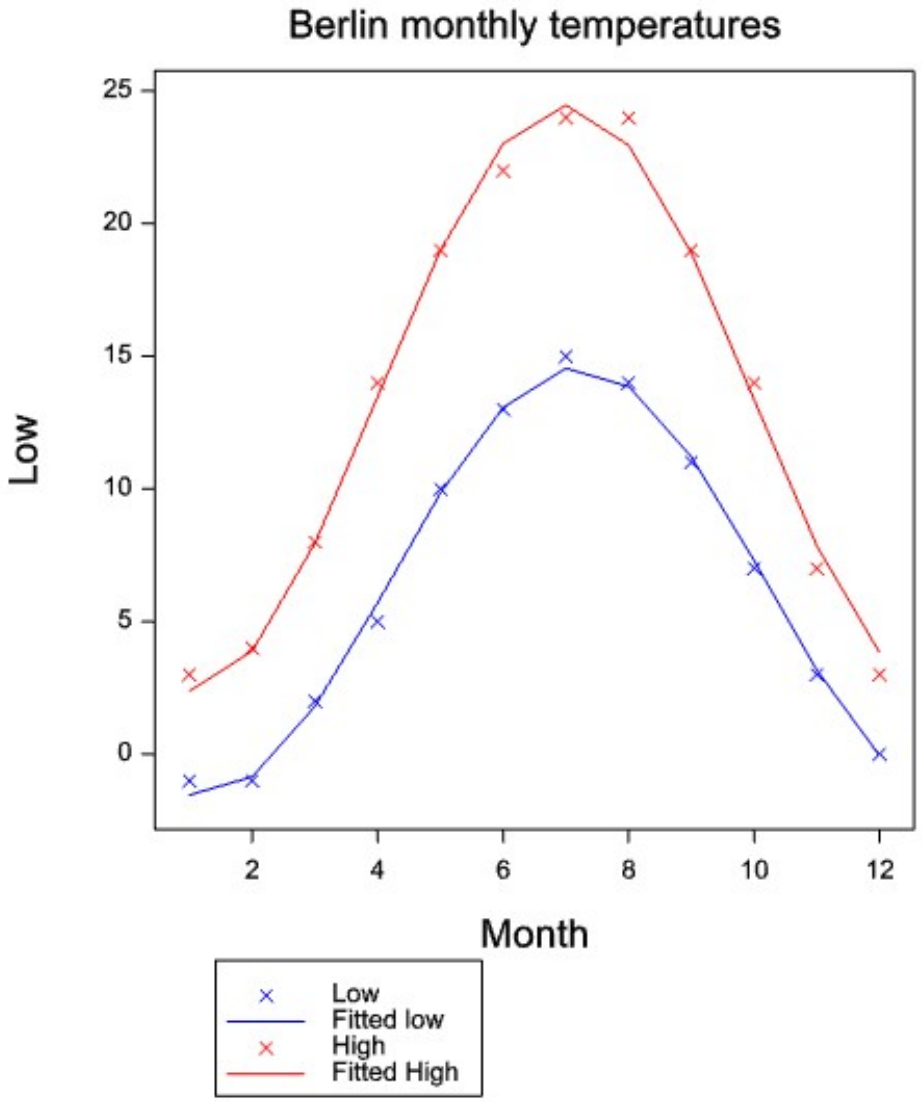


In Summary

- Stratification corresponds to a model in which interactions are fitted also
- It may be that these are not really relevant
- It may be more efficient simply to adjust for the main effect
- This is what ANCOVA does

Continuous predictors

- Forming groups using cut points is a bad idea
- Better is to do some modelling
  - Orthogonal polynomials
  - Splines
  - Fractional polynomials
    - (−2, − 1, − 0.5, 0, 0.5, 1, 2, 3) where the power 0 refers to log
- Sometimes fairly simple modesl can work surprisingly well
  - An example using Fourier analysis of temperatures in Berlin follows



$$temp = \alpha + \beta_1 \cos\left(month\right) + \beta_2 \sin\left(month\right)$$

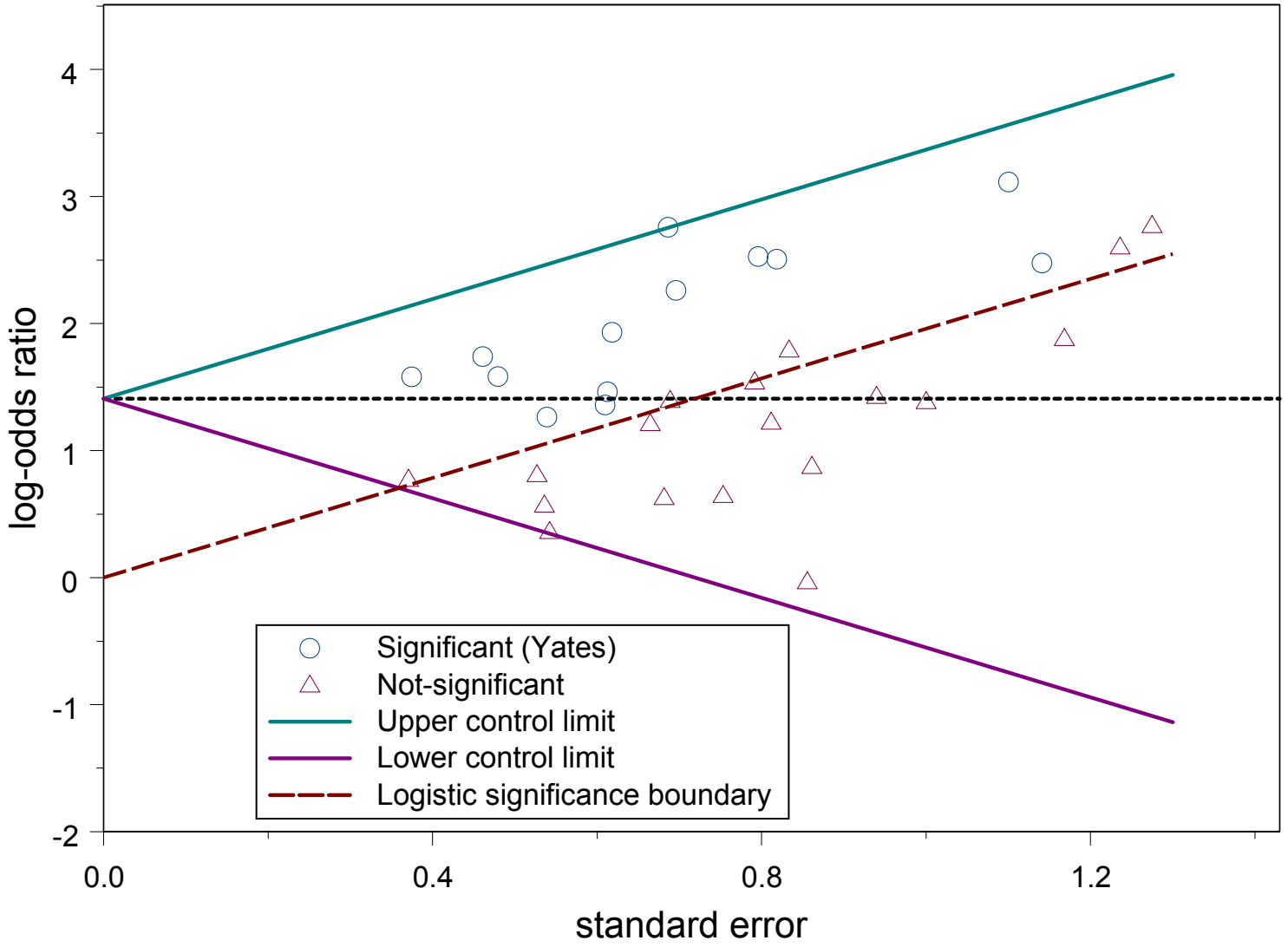
How Many Covariates?

- This is a difficult issue to decide
- In principle (Bayesian perspective) there should be no limit
  - However this only works if you have informative priors
- In practice a few will often bring considerable gains in precision

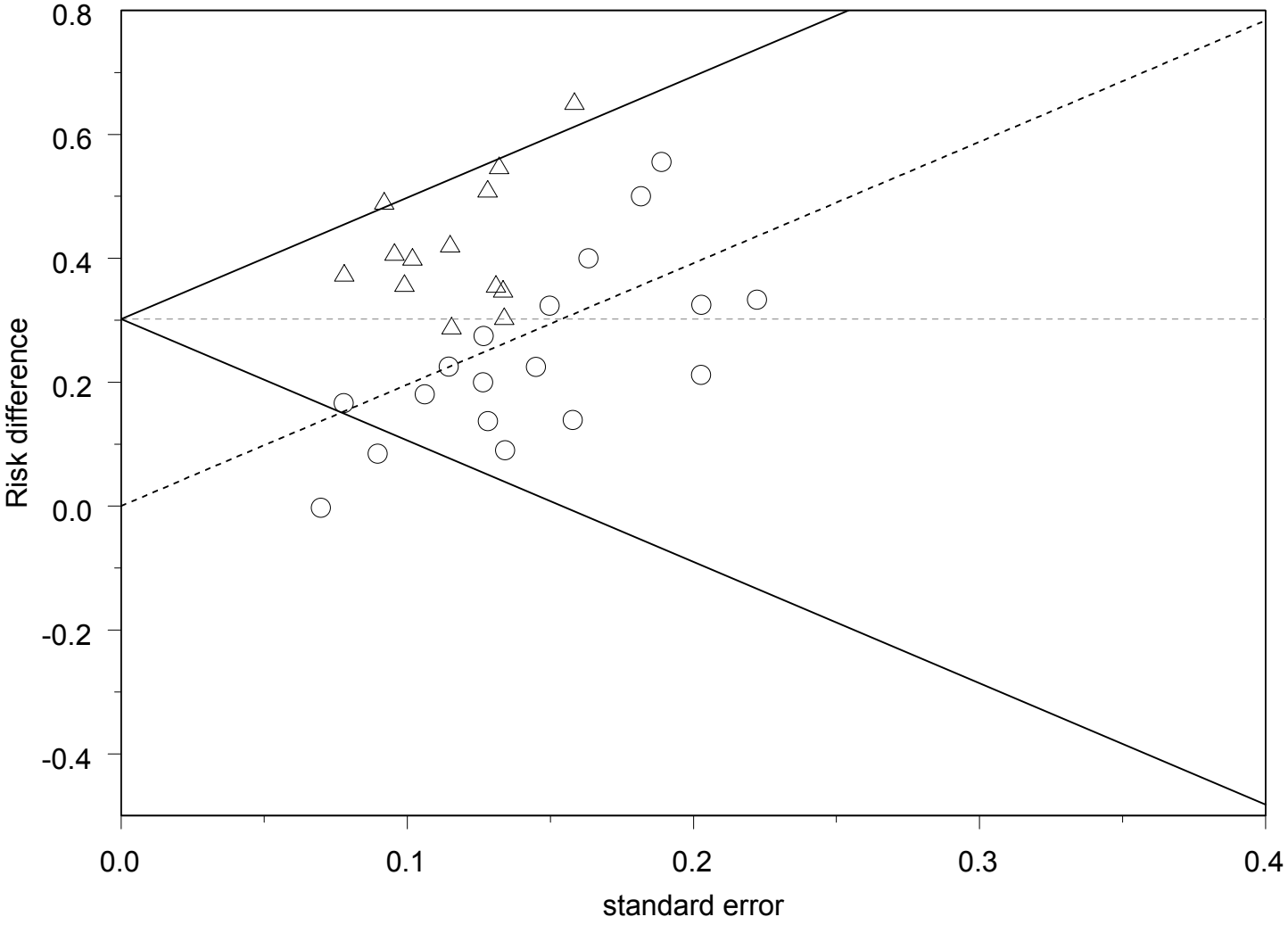
Additive Models

- Finding a good scale for analysis often allows you to use fairly simple models
- A remarkable successful analysis of trials of ulcer healing illustrates the point
- These are of Dan Moerman's data set of 31 trials in cimetidine

31 Placebo-Controlled Trials of Cimetidine



Same Trials on Risk Difference Scale



Moral

- By using the logit scale you can explain everything using a very simple model
- The treatment effect is plausibly constant from trial to trial on this scale
- If, however, you use the risk difference scale then the effect varies from trial to trial
- To model it on this scale you need to use background risk as a covariate

Effect Modifiers

- These are interactions
- They imply that the effect of treatment varies as the value of the covariate varies
- Essentially a form of subgroup analysis is needed
- This is unpleasant and one should try to avoid it if possible by judicious choice of scale
- This is not always possible

Some Principles

- Higher order effects not to be included unless lower order effects in model
- Significance of a factor to be judged by difference it makes to model
- Ideally the covariate main effect model should be established first
- Other things being equal prefer simpler models
- Treatment by covariate interactions should be treated warily unless expected (confirmed)

An Example

- ATAC trial
  - Arimidex, Tamoxifen, Alone or in Combination
- Breast cancer victims
- Patients classified by Oestrogen (ER) and Progesterone (PgR) status
- A v T significant
  - However the effect was only seen in the ER+/PgR- group
- This is a high order interaction and hence inherently implausible

Treatment		A		C		T	
		Patients	Events	Patients	Events	Patients	Events
ER	PgR						
+	+	1930	191	1875	205	1904	222
	-	451	50	492	102	429	102
-	+	63	17	81	22	76	25
	-	233	66	220	71	250	79

Analysis of deviance for logistic regression for the Arimidex, Tamoxifen, Alone or in Combination data illustrating various modelling principles.

Step	Change	df	Deviance	Approximate Chi pr
1	+ ER	1	178.408	<0.001
2	+ PgR	1	47.313	<0.001
3	– ER	-1	-82.388	<0.001
4	+ ER	1	82.388	<0.001
5	+ ER.PgR	1	8.245	<0.004
6	+ Treatment	2	17.705	<0.001
7	+ ER.Treatment	2	0.854	<0.653
8	+ PgR.Treatment	2	10.669	<0.005
9	– ER.Treatment	-2	-5.032	<0.081
10	+ ER.Treatment	2	5.032	<0.081
11	+ ER.PgR.Treatment	2	3.301	<0.192
Total	11	266.494		

df = degrees of freedom; ER = oestrogen receptor; PgR = progesterone receptor.

Fitted and observed events from final model including ER and PgR interaction with each other as well as main effect of treatment but only the treatment by PgR interaction

ER		+		-	
		Expected	Events	Expected	Events
Treatment	PgR				
A	+	191	191	17	17
	-	61	50	55	66
C	+	203	205	24	22
	-	100	102	73	71
T	+	224	222	23	25
	-	93	102	88	79

In Summary

- Modelling is good
- Modelling helps to explain what we see in clinical trials
- Most of the concerns of the regulator can be met by pre-specifying models
- Usually we have far more experience with covariates than with treatments so there is no reason not to make a good job at pre-specifying the covariate model